Title
Brittle smiles: positive biases toward stigmatized and outgroup targets.

Journal
Journal of experimental psychology. General, 142(3)

ISSN
0096-3445

Authors
Mendes, Wendy Berry
Koslov, Katrina

Publication Date
2013-08-01

DOI
10.1037/a0029663

Peer reviewed

Journal of Experimental Psychology: General

Brittle Smiles: Positive Biases Toward Stigmatized and Outgroup Targets

Wendy Berry Mendes and Katrina Koslov

Online First Publication, August 13, 2012. doi: 10.1037/a0029663

Journal of Experimental Psychology: General

0096-3445/12/$12.00 DOI: 10.1037/a0029663

Brittle Smiles: Positive Biases Toward Stigmatized and Outgroup Targets

Wendy Berry Mendes and Katrina Koslov
University of California, San Francisco

We examined individuals' tendencies to exaggerate their positive responses toward stigmatized others (i.e., overcorrect) and explored how overcorrection, because of its fragile nature, could be disrupted. The first 2 studies demonstrate overcorrection: White participants paired with Black partners (Experiment 1A) smiled, laughed, and showed more positive behavior than those paired with same-race partners. Experiment 1B replicated the general effect with a physically stigmatized sample (i.e., facial birthmarks) and then demonstrated that overcorrection is moderated by bias; participants who exhibited more positive behavior toward their partner showed the most physiological "threat" during a stressful task with their partner. We then examined the idea that if overcorrection requires cognitive resources and is effortful, then it may be fragile when resources are taxed. In Experiments 2 and 3, we observed that overcorrection was easily disrupted when resources were compromised (e.g., with stress or cognitive load). Taken together, these studies suggest that positive biases toward stigmatized and outgroup members are fragile and can be undermined when resources are taxed.

Keywords: racial bias, stigma, overcorrecting, psychophysiology

> One sprinkles the most sugar where the tart is burnt.
>
> —Dutch Proverb

In our daily lives, we often have to censor our public face by monitoring our behaviors, expressions, and words, and the ability to do this well is viewed as the foundation of self-regulatory control (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998). We try not to giggle in church, fall asleep during lectures, or ram our car into the jerk who cut us off. However, these control strategies require effort and conscious monitoring, making them potentially vulnerable when resources are taxed. In this investigation, we explored individuals' tendencies to exaggerate their positive behaviors toward and preferences for stigmatized and outgroup members, and then examined how these *correction* strategies, because they are effortful and require resources, can be disrupted with stress or cognitive load.

To examine the processes underlying correction, we focused on racial and stigma domains. These domains are well suited for an investigation that assumes individuals have automatic, uncontrolled responses that are negatively toned and there exists cultural pressures to behave positively. As evidence of the former, in the past decade much evidence has accumulated that observing or interacting with racial outgroup members is associated with sustained fear-conditioning, distress-related cardiovascular responses, and neural activation suggesting fear and uncertainty (Hart et al., 2000; Mendes, Blascovich, Lickel, & Hunter, 2002; Olsson, Ebert, Banaji, & Phelps, 2005; Richeson & Shelton, 2003). These physiological and neural responses are considered automatic or reflexive and certainly not within one's ability to easily control or influence.

A very different story of interracial interactions emerges when examining individuals' controlled responses, specifically what they self-report. Instead of manifestations of threat, fear, and anxiety, White participants appear egalitarian or even lean toward preferring outgroup and stigmatized members (see Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Plant & Devine, 1998). For example, when Whites provided evaluative feedback to an author of a poorly written essay, they rated the content of the essay more positively if they thought the author was Black than when they believed the author was White (Harber, 1998). Following face-to-face interactions, White participants reported greater liking and ascribed more positive traits to stigmatized or different-race partners than to nonstigmatized or same-race partners (e.g., Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Blascovich, Mendes, & Seery, 2002; Mendes et al., 2002). Similarly, Vanman, Paul, Ito, and Miller (1997) showed that even though White participants exhibited facial muscle movements indicating more negative affect while evaluating African American targets, they explicitly reported more liking for those targets. Outside the laboratory, large-scale studies of over 2.5 million respondents found that White Americans consistently self-report less explicit racial bias than they show implicitly (Nosek et al., 2007). These findings suggest that when people can consciously control and monitor their responses, they attempt to correct for racial bias.

Controlled and Automatic Responses in Intergroup Settings

Several research programs have implicated the roles of automatic (implicit) and controlled (explicit) strategies to understand

Wendy Berry Mendes and Katrina Koslov, Department of Psychiatry, University of California, San Francisco.

The research was funded in part by National Heart, Lung, and Blood Institute Grant RO1 HL079383. We gratefully acknowledge Modupe Akinola for her assistance with Experiment 1A, Jim Blascovich for his contribution to Experiment 1B, and Yarrow Dunham and the Health and Psychophysiology Lab summer interns for their assistance in developing and testing the stimuli for the celebrity choice task.

Correspondence concerning this article should be addressed to Wendy Berry Mendes, Department of Psychiatry, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, CA 94143. E-mail: wendy.mendes@ucsf.edu

intergroup behavior. For example, Gilbert and Hixon (1991) showed that participants who interacted with an Asian American experimenter while under cognitive load were more likely to use word completions that were stereotypical (e.g., rice instead of race) compared with those not under cognitive load. One interpretation of this finding is that those under cognitive load could no longer monitor their automatic associations and thus used stereotypical words. In contrast, when not under load, participants might have deliberately steered away from using stereotypical words in the presence of a racial minority because they did not want to appear insensitive or be perceived as racist. Similarly, Devine (1989) concluded that regardless of levels of explicit prejudice, White individuals have knowledge of cultural stereotypes, and when participants are unable to consciously monitor their behavior, they use stereotypes to make judgments of others. Only when cognitive resources were available did individuals with lower explicit prejudice replace the stereotyped thoughts with more egalitarian responses.

Another model of dual processes in response to stigma, developed by Pryor, Reeder, Yeadon, and Hesson-McInnis (2004), took into account the temporal displacement of automatic and controlled responses to stigmatized others. They posited that one's initial reaction to stigmatized others is via the reflexive or associative system, and that a reflective, rule-based system only has the opportunity to make adjustments to this initial reflex. For example, they found that individuals high in disgust sensitivity were more likely to keep their distance from stigmatized targets, such as an individual with AIDS, in the initial seconds after target exposure, indicating their lack of positivity for the target, but they would move closer over time, presumably to correct for their initial responses.

These correction efforts might stem from genuine desires to be egalitarian or from impression-management concerns not to appear prejudiced (Devine, Monteith, Zuwerink, & Elliot, 1991; Norton, Sommers, Apfelbaum, Pura, & Ariely, 2006). Given the discrepancy between automatic responses and controlled reactions, we argue that in many cases, those who correct may be the ones who are the most biased toward outgroup or stigmatized group members. Some data support this hypothesis. High-prejudice participants, who were concerned about whether a minority group partner might perceive them as prejudiced, engaged in more intimacy-building behaviors—such as increased eye contact, more responsiveness, and positive regard—during a video message to their partners (Vorauer & Turpie, 2004). These corrective measures, however, may take their toll on cognitive resources. Richeson et al. (2003) found that high-prejudice Whites who interacted with minority group members made more controlled processing errors after the interaction, suggesting that interracial interactions require more conscious control and can deplete resources in a subsequent cognitive task.

What these models share is an acknowledgment that there are multiple systems that respond to the presence of stigmatized or minority group members, and that one's automatic responses can be in conflict with the explicit behavior one wishes to display. These models also recognize that engaging in unprejudiced behavior may require resources (e.g., executive control, processing time), a motivation to appear unprejudiced, and situations in which either norms of egalitarianism are activated or responses are publicly expressed.

Evaluative Concerns

A critical moderator of these proposed effects is the extent to which behavior can be evaluated or judged. When evaluative concerns are low or cultural norms are aligned with implicit attitudes, there is little reason to monitor or correct behavior; yet when evaluative concerns are high or behavior is public, individuals' tendency to correct their racial bias is expected to be elevated. Indeed, past research has shown the importance of public expression and evaluative concerns on intergroup behavior.

In an early demonstration of the importance of evaluative concerns on intergroup treatment, Dutton and Lake (1973) examined how much money participants gave a Black compared to a White panhandler, following a manipulation in which participants were led to believe they had higher levels of racial bias (or not), using false feedback from physiological signals. White participants who believed they were implicitly prejudiced gave more money to Black panhandlers than those who were not led to believe they were prejudiced or those confronted with a White panhandler. Similarly, Vorauer and Turpie (2004) found that when individuals were under pressure to appear unprejudiced, those who were higher in racial bias managed to "shine"—to behave in a highly positive manner toward outgroup members. In contrast, those lower in bias, when faced with the pressure to appear unprejudiced, became overvigilant of their own responses and "choked" under the intergroup pressure. These evaluative concerns may engender an "identity threat." In a recent study by Harber, Stafford, and Kennedy (2010), participants whose egalitarian ideals were threatened prior to evaluating an essay provided less criticism and more positive feedback to authors they thought were Black compared with those they thought were White, suggesting that individuals may engage in increased monitoring of their behavior under a threat to their egalitarianism.

One potential by-product of evaluative concerns is that individuals may attempt to appear "colorblind" in their interactions with people from other racial groups (Apfelbaum, Sommers, & Norton, 2008). For example, White participants avoided using an efficient strategy to narrow down a list of possible individuals if the strategy simply required indicating a person by their (minority) racial category. Importantly, colorblindness can be counterproductive, as participants who behaved in a colorblind manner toward Black partners were judged as less friendly, and the process of inhibiting themselves depleted their cognitive resources.

Resource-Dependent Model of Intergroup Behavior

Taken together, these research streams suggest that interacting with stigmatized or outgroup members may be viewed as the quintessential self-regulatory domain, since these situations require monitoring one's behavior and applying corrective measures as needed. However, the capacity for self-regulation is a limited and consumable resource, such that one act of self-control reduces one's subsequent self-control (Baumeister et al., 1998). Self-regulation is thus vulnerable to fatigue, and as demands on the executive self increase, one is more likely to experience failures of self-regulation: inability to remain vigilant about the status of one's goals and to inhibit one's prepotent responses. If appearing unprejudiced is an act of self-regulation for some individuals, then by exhausting their resources, they may be less likely to engage in corrective behavior.

Figure 1 presents a resource-dependent model that integrates automatic and controlled attitudes, evaluative concerns (via norms), and emotional and cognitive resources (cf. Trawalter, Richeson, & Shelton, 2009). In this model, we propose that evaluations and treatment of any target object or person is a result of automatic or reflexive associations and conscious, deliberate evaluations of that target. When automatic associations and controlled evaluations are in agreement (either positive or negative), the treatment or behavior directed at the target is simply valenced in line with the evaluations. However, when there is a discrepancy between the controlled and automatic attitudes, several critical factors can influence subsequent behavior. First, the context is important; when responses are made that can be evaluated (currently or in the future), individuals will monitor their responses and apply corrective behavior as needed. When future evaluation by self or others is unlikely, monitoring is low, and correction is unlikely to occur. In the current research, we only consider situations in which evaluative concerns are high.

Assuming monitoring of responses is activated, the second critical feature of this model is the extent to which socioemotional resources are high or compromised in some way. Socioemotional resources include executive ability to actively regulate emotions, thoughts, and behaviors, and can be compromised by high emotional arousal, cognitive load, exhaustion, and time pressure or with drugs and alcohol (e.g., Bartholow, Dickter, & Sestir, 2006).

The ability to inhibit automatic responses requires high socioemotional resources. When resources are high, we expect individuals to correct for their outgroup biases; however, when resources are low, we expect that correcting biases is more difficult. Specifically, if a target person is implicitly evaluated negatively and resources are high, the negative attitudes can be corrected, and we expect positive treatment and behavior toward the target person. In contrast, when implicit attitudes are negative and resources are low, we expect correction efforts to fail and behavior to be valenced as the reflexive component—that is, we expect less positive treatment.

The Current Research

We integrated self-regulation and intergroup threat theories to arrive at three predictions. First, we reasoned that Whites, and not Blacks, would be more likely to be concerned about appearing racially prejudiced, so we predicted that in the context of a casual (i.e., nonstressful) interaction, White participants would engage in more positive expressions and behavior when interacting with Black partners than Black participants with White partners or same-race dyads. Second, we predicted that implicit bias would moderate these responses such that those who exhibited the most bias would also be the individuals showing the most positive behavior during a casual interaction. In this case, we operationalized bias as physiological responses during a face-to-face social interaction with stigmatized (or nonstigmatized) partners (Experiments 1B and 3) or implicit racial bias (Experiment 2). Finally, to establish that overcorrection is fragile and requires resources, we examined intergroup choices following a resource depletion manipulation. In Experiments 2 and 3, we predicted that resource depletion would disrupt people's ability to overcorrect. In Experiment 2, we examined the effects of challenge (resources intact) versus threat (resources compromised) states on intergroup choices, and in Experiment 3, we manipulated cognitive load versus no load during an intergroup choice task.

Experiment 1A

Method

Participants. White and Black female undergraduates (n = 64; 52% Black) between the ages of 18 and 23 years (M = 21,

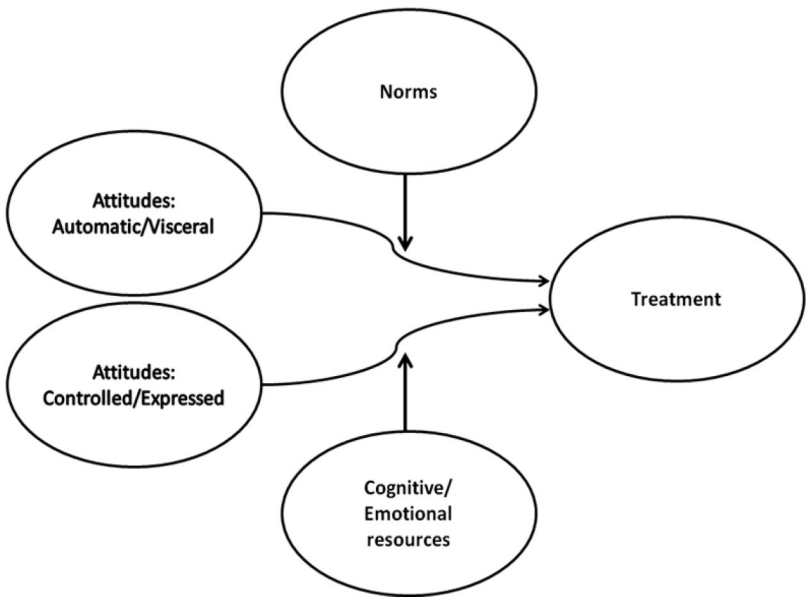

Figure 1. Resource-dependent model of intergroup behavior.

SD = 2) were recruited from colleges and universities around the greater Boston area for a study described as "physiological responses during laboratory tasks."

Procedure. Participants arrived at the laboratory and were randomly assigned to interact with either a White or Black female confederate. Confederates (N = 10; 50% Black) were trained together to optimize similarity, and their behavior and affect were carefully scripted and monitored so that we could interpret the behavior of the participants independent of how the confederate acted. During this initial interaction, participants were told that they would be sharing information about themselves by asking each other questions. We provided the questions to the dyad, which were constructed so that the content was relatively neutral and nonthreatening to disclose to a stranger, but would offer some initial insight into the person (e.g., "What type of websites do you frequently visit?"). The social interactions took place with the dyad sitting across from each other, and the interaction was videotaped from a dual camera perspective with split-screen capabilities so we could code participants' behavior as well as monitor confederates' behavior for adherence to the protocol. Here we focus on the first 2 min of this initial interaction (Ambady & Rosenthal, 1992), which we coded specifically to answer the questions posed here. As Ambady and Rosenthal (1992) have shown, the predictive accuracy of expressive behavior can be identified in time sequences as short as 0.5–4 min.

Videocoding. We identified a priori observable positive behavior: frequency and duration of participants' smiles, frequency of nodding and laughing, how happy participants appeared, how many expressed positive statements were directed toward their partner (e.g., "that's great!" "awesome"), and how much the participants appeared to like their partner. The 2-min interaction was coded in 10-s intervals. Frequency measures were coded as either present or absent at any point during the 10-s interval (possible range from 0 to 12). The nonfrequency measures were estimated after watching the full 2-min video interaction. Smiling duration was estimated between 0% and 100% of the time, and the remaining questions (expressed attitude, liking, and happiness) were rated from −3 to 3 with appropriate anchors and neutral (0) points indicated.

We trained three female coders (two White, one African American), who independently scored each social interaction, focusing only on the participant and blind to the confederate's race (only the participant was visible during the videocoding). Alphas across these judges were good (.71–.89).

Results

To test our predictions that (a) intergroup interactions would result in significantly more positive expression and overt positive behavior and (b) White participants would express more positive emotions and behaviors when interacting with Black confederates than vice versa, we conducted a series of analyses of variance examining the race of the participant, the intergroup context (same-race or different-race partner), and their interaction.

We observed main effects for participants' race for frequency of smiling, $F(1, 62) = 26.13$, $p < .001$; laughing, $F(1, 62) = 9.16$, $p < .004$; and nodding, $F(1, 61) = 18.02$, $p < .001$. On average, White participants smiled, laughed, and nodded more frequently during the first 2 min of the interaction than Black participants. No other main effects were significant.

These main effects, however, were qualified by significant interactions. Indeed, for every variable, except nodding, we observed the anticipated two-way interaction (see Table 1). Consistent with our predictions, the interactions were driven by White participants paired with Black partners (intergroup context) who smiled more often and of longer duration, laughed more, appeared to like their partner more, expressed more positive statements, and appeared happier than participants in the other three conditions. Planned contrasts focusing on just intergroup dyads revealed that all of the positive behavior (including nodding, but not liking) occurred significantly more among White participants interacting with Black partners than Black participants interacting with White partners. These data strongly supported our predictions that although intergroup interactions might result in more superficial positive emotions and behaviors, it seems to be isolated to White participants acting more positively toward Black partners, rather than Black participants showing increased positive behavior toward White partners.

Experiment 1B

In Experiment 1B, we were interested in replicating and generalizing the overcorrection effect to another stigmatized group that individuals might feel pressure to treat more positively. In addi-

Table 1
Summary of Positive Expressions During Social Interactions by Race of Participant and Racial Composition of Dyad

	Intragroup				Intergroup				
	Black		White		Black		White		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Interaction *F*
Smiling frequency	6.0_{ab}	2.6	7.1_{b}	1.6	4.7_{a}	2.6	9.2_{c}	1.8	9.01^{**}
Smiling duration (% time)	41_{a}	25	38_{a}	17	37_{a}	29	60_{b}	21	4.56^{*}
Laughing	1.7_{a}	1.5	1.2_{a}	1.4	1.3_{a}	1.2	3.3_{b}	2.3	4.19^{*}
Perceived liking	0.8_{b}	1.1	0.3_{a}	0.5	0.5_{ab}	1.0	0.8_{b}	0.6	4.03^{*}
Positive statements	0.8_{ab}	1.0	0.6_{a}	0.5	0.5_{a}	1.3	1.3_{b}	0.8	4.47^{*}
Nodding	2.7_{a}	2.2	5.5_{b}	1.7	3.1_{a}	2.1	5.3_{b}	3.0	0.33
Appeared happy	0.8_{ab}	1.0	0.6_{a}	0.7	0.5_{a}	1.2	1.5_{b}	1.2	5.18^{*}

Note. N = 64. Different subscripts indicate significant post hoc differences within a row.
* $p < 05$. ** $p < .01$.

tion, we examined whether participants who displayed more positive behavior during a casual exchange would be those who were higher in bias toward that stigmatized group member. In this study we operationalized racial bias using cardiovascular responses, during a stressful, cooperative task with the same partner (e.g., Dovidio, Pearson, & Orr, 2008). We were interested in testing the hypothesis that bias would be positively related to friendly treatment of stigmatized group members. To increase generalizability of the finding, we chose a stigmatizing characteristic that was not based on race or ethnicity but was still perceived as relatively essential to the person: We manipulated whether the partner had a physical stigma (facial birthmark). We used data we had previously collected that manipulated whether a female confederate bore a port-wine stain facial birthmark created with makeup (see Blascovich et al., 2001, Experiment 2). Previously, we published data showing that female participants interacting with a stigmatized partner exhibited a malignant pattern of cardiovascular reactivity (defined as inefficient cardiac output and increased vascular resistance). For this investigation we coded the videotapes to examine whether positive behavior during a casual initial interaction would predict physiological responses from a stressful task with the same partner.

Method

Participants. Female participants ($N = 39$) were recruited from a large university campus. All participants were prescreened for heart murmurs, pregnancy, and use of medication affecting cardiovascular responses.

Stigma manipulation. Female confederates arrived at the laboratory 30 min prior to the start of the study to prepare. The experimenter determined condition from random assignment and without telling the confederate what the assignment was; the confederate sat down and closed her eyes for the duration of makeup application. In the stigma condition, the experimenter applied theatrical makeup to form the appearance of a large port-wine stain birthmark on the right cheek of the confederate. In the control condition, similar brushes and applications were applied to the face, but only translucent powder was used, so that at the end of the makeup application the face was clean of any marks. This process assured us that the confederate would be unaware of experimental condition. Once the makeup was complete, the confederate went outside the laboratory to wait in the hallway for the participant to arrive.

Procedure. Participants arrived at the laboratory and were introduced to the confederate. After informed consent was completed, the dyad sat in the same room in adjacent chairs and was instructed to "get to know each other." We provided a sheet to guide them during this interaction that prompted them to share their age, hometown, major, and other basic demographics. The participant was instructed to begin, and the confederate provided a scripted background that was based on a composite of the "average" female student at this university. The interaction lasted between 2 and 3 min, and we videotaped the interaction.

The dyad was then separated, and physiological sensors were applied. This included impedance cardiography, which estimates cardiac output (CO; the amount of blood ejected from the heart in 1 min), and continuous blood pressure data, which were used to calculate total peripheral resistance (TPR), the amount of vasoconstriction versus dilation occurring in the arterioles (see Mendes, 2009, for details). Challenge (benign) states tend to result in increased CO and decreased TPR, whereas threat (malignant) states typically result in decreased CO and increased TPR. After a resting baseline period, the dyad then interacted with each other during cooperative, stressful tasks in which we measured cardiovascular changes.

Videocoding. We again coded the videotaped social interaction, looking for positive behavior directed toward the participant's partner, specifically smiling. We focused on frequency of smiling and again determined whether any smiling was present or absent at least once during every 10-s interval of the first 2 min of the information exchange. Six research assistants coded at least 25% ($n = 9$) of all participants, and every participant was coded by at least three coders. Reliability was high across coders ($\alpha = .87$). Six videos could not be coded because of either loss of audiovisual connection during the interaction or incompleteness.

Results

As in Experiment 1A, participants paired with stigmatized partners (i.e., those with a facial birthmark) smiled more often ($M = 5.9$) than participants assigned to nonstigmatized partners ($M = 4.2$), $F(1, 31) = 4.27$, $p < .05$. We then examined cardiovascular responses[1] from the cooperative social interaction to determine whether physiological changes were related to smiling behavior. We predicted smiling scores using a composite of cardiovascular reactivity (z-transformed TPR reactivity and CO reactivity combined into a single score such that higher scores indicated more "threat"), the partner's stigmatized status (0 = no birthmark, 1 = birthmark), and the interaction of cardiovascular threat and partner's stigma. The interaction was significant, $t(29) = 3.42$, $p < .01$. As can be seen in Figure 2, among participants interacting with nonstigmatized partners, there was a negative relationship (though short of significant) between smiling and threat responses ($\beta = -.25$, $p < .10$), with greater smiling associated with lower threat reactivity when participants were paired with nonstigmatized partners. In contrast, among participants interacting with stigmatized partners, the relationship between smiling frequency and physiological threat was significant and positive ($\beta = .63$, $p < .01$). Supportive of the biased-based correction hypothesis, participants who were highest in bias (measured with cardiovascular reactivity) were more likely to smile during a casual, social interaction with their stigmatized partner.

Discussion: Experiments 1A and 1B

Experiment 1A provided evidence that White participants paired with Black partners engaged in more observable positive behavior than same-race dyads or Black participants paired with White partners. This is not surprising given that even though Whites and Blacks show similar physiological patterns in different-race interactions relative to same-race interactions (e.g., Mendes, Major, McCoy, & Blascovich, 2008), there are divergent goals in interracial interactions (Bergsieker, Shelton, & Richeson, 2010); whereas Whites strive to be liked, Blacks desire respect. The

[1] Three participants did not have blood pressure data for this period and were not included in this analysis.

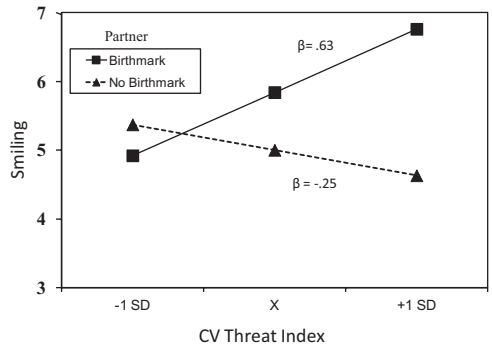

Figure 2. Experiment 1B: Relationship between smiling and cardiovascular (CV) reactivity by partner's stigmatized status.

increased desire to be liked, identified by Bergsieker et al. (2010), nicely aligns with the results of Experiment 1A, which show that Whites and not Blacks engage in more positive behavior during interracial interactions.

Experiment 1B replicated the effect of more positive treatment toward stigmatized partners, this time with physically stigmatized partners. In addition, we showed support for the idea that positive behavior toward stigmatized partners was associated with individuals' bias: Those who exhibited more malignant stress reactivity during a stressful cooperative task with their partner were more likely to show positive behavior during the initial interaction. In contrast, more smiling was related to less threat (though not significant) among participants paired with nonstigmatized partners. Our assumption is that the smiles we coded were likely qualitatively different toward stigmatized compared with nonstigmatized partners (see Niedenthal, Mermillod, Maringer, & Hess, 2010). Though we lacked the video resolution and adequate close-ups to properly code for genuine (Duchenne) smiles, we believe that the differential relationships between smiling and cardiovascular reactivity suggest that smiling toward stigmatized partners was, in part, indicative of more anxiety or uncertainty.

After establishing these effects, we explored the idea that if overcorrection is effortful and requires cognitive resources, then it might be easily undermined when resources are taxed. To test this idea, we reduced the ability to engage in these corrective measures by placing participants in highly stressful situations or by taxing their resources with cognitive load. We also developed racial preference tasks, with the goal that participants could monitor and control their behavior on these tasks if they had the cognitive resources to do so, but when resources were taxed, it would be more difficult to keep track of their expressed preferences.

In Experiment 2, we exposed all participants to a stressful task that would reliably increase sympathetic arousal. Based on prior research, however, the profile of arousal changes could be further differentiated into threat or challenge states: Some participants would experience the task as threatening, whereas others would experience the task as challenging. Threat states are argued to occur when demands outweigh resources, whereas challenge states occur when resources outweigh demands. Therefore, we predicted that participants for whom the stressful task resulted in threat responses would have exhausted their resources and would no longer be able to engage in corrective strategies. Given the predictions that corrective strategies are more likely to be used among those higher in racial bias, we expected that higher racial bias participants (measured with the Implicit Association Test [IAT]) would correct their racial preferences when they had the resources to do so (i.e., when they were challenged). In contrast, higher racial bias individuals who were threatened and would not have the necessary resources to correct their racial bias tendencies would not show corrective behavior. Among lower racial bias individuals, we did not expect challenge or threat profiles to influence their racial preferences.

Experiment 2

Method

Participants. White participants ($N = 61$; 67% female; $M_{age} = 21.8$ years, $SD = 3.3$) were recruited to participate in a psychophysiology study. Exclusion criteria included the same factors as Experiment 1B.

Procedure. Upon arrival at the laboratory, participants completed an IAT (Greenwald, McGhee, & Schwarz, 1998) to provide estimates of individuals' implicit racial bias. Following this, we applied sensors to record cardiovascular changes (described below). After a 5-min resting baseline, participants were instructed to prepare and deliver a speech to a panel of evaluators. Upon consenting to this task, two White research assistants (one male and one female) entered the room and provided more details of the speech task. Specifically, participants were told that they would be delivering a 5-min speech on why they should be hired for their dream job. The participants were then left alone to prepare for 5 min, after which the interviewers reentered the room, sat across from the participants, and instructed the participants to begin. The evaluators also asked scripted opinion questions (e.g., "Is it true that the customer is always right?"). Immediately after the interview task, participants were presented with the race preference task, which was described as a *résumé selection task*.

Résumé task. On the basis of Bertrand and Mullainathan's (2003) study of employability according to name of job candidate, we created a set of 24 résumés. Six of the résumés were designed to be high quality, 12 were medium quality, and six were low quality. Quality was manipulated by differing educational background, work experience, and work skills. A panel of 11 pretesters rated each numbered—but unnamed—résumé on a −4 to 4 scale for education, work experience, work skills, and overall quality; ratings on these four variables were highly correlated ($\alpha = .97$), so we averaged the four to create a "quality" variable. An analysis of variance confirmed that our raters differentiated between low-, medium-, and high-quality résumés, $F(2, 21) = 73.49$, $p < .001$. The mean ratings were 3.37 ($SD = 0.18$) for high, 2.08 ($SD = 0.37$) for medium, and 0.29 ($SD = 0.70$) for low. Planned contrasts between the mean quality ratings of each group were significant at the $p < .001$ level. We then assigned two thirds of the résumés stereotypically Caucasian first names (e.g., Jake, Bradley, Emma, Caitlin) and the other third stereotypically African American first names (e.g., Darnell, Trevon, Shanice, Aaliyah), all paired with the most common American last names (e.g., Moore, Young, Lewis, Williams).

After the interview task, participants were given the résumé task to complete. They were instructed to screen the résumés for

invitations to a first-round interview for an executive management position at a Fortune 500 company and were asked to select eight résumés for consideration. We instructed participants that they would receive a $5 bonus if they successfully chose the best eight candidates from the set. The experimenter timed the task with a visible stopwatch, and participants received 5 min to complete the task.

Physiological measures. We used impedance cardiography, electrocardiography, and blood pressure measures to calculate CO and TPR. To obtain impedance cardiography signals, we applied four Mylar bands around the neck and torso, which were then attached to a HIC-2000. Electrocardiography was obtained with a standard Lead II configuration (right arm, left leg). Signals were obtained at 1,000 Hz and integrated with an MP150 hardware system. Blood pressure was obtained with a tonometric blood pressure monitor (Colin CBM-7000; San Antonio, TX) attached to participants' nondominant arms (see Mendes, 2009, for additional details). Signals were edited and scored offline by trained research assistants using Mindware software (heart rate variability, impedance cardiography, and blood pressure modules).

Results

We first tested whether the interview task reliably increased sympathetic activation from baseline. Participants showed significantly increased heart rate ($M = 16.6$), $t(60) = 14.07$, $p < .001$, and decreased preejection period (a measure of sympathetic activation; $M = -7.31$), $t(60) = -7.75$, $p < .001$. We then examined changes in CO and TPR during the interview and observed significant variability: About 60% of the participants showed a challenge profile (decreased TPR and increased CO), whereas the remaining participants showed a threat profile (increased TPR and decreased CO). We combined these measures into a single index in which higher numbers indicated more threat and lower numbers indicated more challenge.

Before testing our main hypothesis, we examined the relationship between IAT racial bias and the cardiovascular threat index. As the stressor was not race relevant—that is, the interviewers were the same race as the participants, and there was no interview content that was race related—we assumed these two predictors would be uncorrelated, and they were ($r = .04$, *ns*). Given that the two primary predictors were uncorrelated, we used them as independent predictors of race preference from the résumé selection task.

To test our main hypothesis, we used centered IAT and the threat index to predict the number of medium-quality African American résumés chosen (we examined choice of medium-quality résumés because this was where there was ambiguity regarding who should be chosen). Given our outcome variable had a limited range (0–3), we used logistic regression[2] in which we created a dichotomous outcome such that a code of 0 represented either 0 or 1 medium-quality Black candidate résumés chosen and a code of 1 was assigned to cases in which two or three medium-quality Black résumés were chosen. It is important to note that we asked for eight résumés to be selected; with only six high-quality résumés present, at least two medium-quality résumés must be selected.[3] If an individual was colorblind or attempting to engage in affirmative action decisions, the number of medium-quality Black résumés chosen would be 0 or 1, given that only one third of the applicants were African American. To choose two (or more) medium-quality Black résumés would be to ignore base rates and suggest overcorrecting.

The hierarchical logistic regression produced a nonsignificant model with the IAT and physiological reactivity as predictors of résumé choice, $\chi^2(2, N = 44) = 3.01$, $p = .22$. In the second step, the addition of the interaction term significantly improved model fit, $\chi^2(1, N = 44) = 7.84$, $p < .006$. Odds ratio estimates of participants high in racial bias and low in threat yielded an odds ratio of 6.7 (95% CI [1.31, 34.16]) of selecting medium-quality Black résumés (see Figure 3). In line with our predictions, the participants who experienced challenge and who had greater implicit racial bias were more likely to overselect medium-quality Black candidates. Because these individuals were not experiencing threat, they presumably still had the cognitive resources that allowed them to correct their preferences. However, among participants experiencing threat, higher racial bias was associated with choosing fewer Black candidates. Among those lower in racial bias, current physiological state was not related to candidate preferences.

Discussion

Experiment 2 revealed that during stressful situations, those higher in racial bias displayed preferences for Black candidates, but only if they had the resources to do so as indicated by their physiological profile. Those higher in racial bias who were coping poorly—as evidenced by greater threat responses—were less likely to select medium-quality Black candidates. These data support the idea that overcorrection requires resources, and when resources are taxed (as indicated by threat), overcorrection is less likely to occur.

Thus far, we have shown that majority group members engage in overcorrection during low-stress situations, those who overcorrect are more likely to be higher in racial bias, and stressful situations that overwhelm resources can reduce overcorrection. Importantly, one limitation of Experiment 2 was that our approach to examine resources was based on individual differences in stress responding, so third-factor explanations cannot be ruled out. In the last experiment, we randomly assigned participants to a cognitive load manipulation to experimentally test the idea that reducing cognitive resources would reduce overcorrection.

Experiment 3

Experiment 3 used a two-experiment ruse, in which participants first completed a *celebrity choice task* either under cognitive load or not (Experiment 3A), and then in a separate, ostensibly unrelated study, participants completed a stressful, evaluative task in the presence of two Black interviewers, during which we estimated their racial bias using cortisol changes during the evaluation (Experiment 3B). Instead of using the IAT as an indication of racial bias, we used cortisol reactivity from a social interaction with

[2] These effects replicate when using multiple regression instead of logistic regression.

[3] It was possible for three medium-quality Black résumés to be selected by sacrificing a high-quality White résumé, which some participants elected to do.

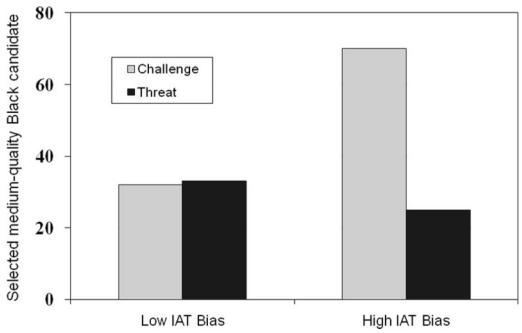

Figure 3. Experiment 2: Probabilities of selecting a medium-quality Black candidate as a function of Implicit Association Test (IAT) racial bias and current threat state.

outgroup members as a proximal measure of intergroup anxiety (Page-Gould, Mendoza-Denton, & Tropp, 2008). We had two predictions. First, participants completing the celebrity choice task under "no load" would prefer equivalent (or greater) numbers of Black than White celebrities, more than those completing the choice task under "load." That is, we expected that individuals would attempt to appear egalitarian and, when not under cognitive load, would have the resources necessary to correct their preferences; under load, decreased resources would make it more difficult to keep track of corrected preferences. Second, increases in cortisol, a proxy for racial bias, would predict greater preferences for Black celebrities for participants not under cognitive load (consistent with the biased-based correction hypothesis), but for participants under cognitive load this effect would be attenuated or reversed.

Method

Setting and participants. Two laboratories in the psychology department at Harvard University served as the experimental setting. We recruited 60 White participants for Experiment 3A, of which 54 completed Experiment 3B. As this experiment included a two-experiment ruse, the time between completing the experiments ranged from hours to weeks. During debriefing after Experiment 3B, no participants suspected that the two studies were related. We examined time between the experiments as a potential predictor of the effects observed, and there were no effects regarding how closely in time participants completed the two experiments.

Participants were recruited through the department's online study pool and excluded for conditions affecting endocrine products, as well as for social anxiety. For completing Part A, they received either a half hour of course credit or $5, and for completing Part B, they received an hour and a half of course credit or $25. Of the 54 participants who completed both experiments, 39 were male and 15 female, and the mean age was 22.63 years ($SD = 6.23$). Twenty-nine participants had completed Part A in the load condition and 25 in the no-load condition. Part B of the experiment was always run in the afternoon to control for diurnal variation in cortisol.

Experiment 3A: Celebrity choice task and cognitive load manipulation. The experimenter began by playing a CD track that had a recording of four computerized musical instruments playing a C note: piano, trombone, flute, and violin (Knowles & Condon, 1999). If participants could not distinguish between the instruments, the experimenter played the four tones again. Next, the experimenter explained the celebrity choice task[4] to the participants, telling them that we were interested in their preferences for various celebrities. We developed this task after extensive pilot testing, during which we equated celebrities (actors, athletes, musicians, and politicians) on four dimensions: likability, familiarity, attractiveness, and media attention. During each trial, the participants were presented with two celebrities, matched on sex, approximate age, and celebrity genre, and were asked to indicate which one they preferred. Out of 60 total trials, 20 critical trials consisted of choosing between a White or Black celebrity.

Participants were told that in each trial two images of celebrities would be presented on the screen, one on the left and one on the right. Participants were told to indicate which celebrity they preferred using the keyboard's *E* and *I* keys, respectively. The experimenter elaborated that there were no wrong answers and that if they did not know one or both of the celebrities, they should go with their gut instincts. Participants were asked to go quickly, taking only a few seconds for each decision.

The experimenter stayed in the room while participants completed 10 practice trials and confirmed that they understood the task. After the participant finished the practice trials, the experimenter explained that she was going to start another track of musical tones. Participants in the cognitive load condition were told that they would have to count the number of piano tones they heard, starting from the beginning of the track and ending when they completed the celebrity choice task. Participants in the control condition were told that the tones were playing as background sounds to mask any distracting environmental noise. All participants were told not to begin the choice task until they were told to do so by the experimenter. The experimenter then started the CD track and a stopwatch and returned to the control room. After 90 s, all participants were told over the intercom to begin the choice task. Participants informed the experimenter when they were finished, and this time was recorded.

Experiment 3B: Interracial stress task. Upon arrival at the laboratory, participants rested for 30 min prior to providing an initial saliva sample. Following this, the experiment was very similar to Experiment 2; we instructed participants that they would complete a mock job interview in front of a panel of evaluators. In this case, however, we were interested in activating stress associated with a racial context, so the interviewers were always two African Americans (one male, one female). The interviewers provided the description of the speech task, left participants alone to prepare, and then reentered the room and sat across from the participants while they delivered the speech. Following the speech, participants were then instructed to complete a serial subtraction task in which they counted backwards by steps of 7 from a four-digit number. After this the interviewers left and independently completed a rating of how anxious the participants appeared. We collected a second saliva sample following the stress task.

Neuroendocrine measures. Neuroendocrine samples were obtained with IBL SaliCap sampling devices, which require par-

[4] The celebrity choice task can be obtained from the authors.

ticipants to expectorate 1 ml of saliva into a cryovial via a plastic straw. Saliva samples were stored immediately at −80° C until they were shipped overnight on dry ice to a laboratory in Dresden, Germany, where they were assayed for salivary-free cortisol. Intra- and interassay coefficients of variance were less than 10%.

Evaluators' ratings. To gauge participants' behavioral reactions, both interviewers independently completed a questionnaire at the end of each interview. They reported how anxious participants appeared during the tasks on a 9-point scale (−4 to 4).

Results

Experiment 3A. We predicted that participants assigned to the no-cognitive-load condition would choose a higher proportion of Black over White celebrities than those under cognitive load. Confirming our prediction, participants assigned to the no-cognitive-load condition chose a higher percentage of Black celebrities (52%) than those assigned to the load condition (42%), $F(1, 58) = 6.30$, $p < .02$. If participants were choosing roughly equal numbers of Black and White celebrities, then the percentage of Black celebrities chosen would be 50%. Indeed, in the no-load condition, participants' choices were no different from 50%, $t(27) = 0.79$, *ns*. In contrast, participants assigned to the load condition chose significantly fewer Black celebrities than would be expected if participants were attempting to respond with no racial bias, $t(31) = -2.84$, $p < .008$.

Experiment 3B. We then tested whether racial bias, measured with cortisol changes from the interracial interaction, could predict celebrity choice and examined the moderating role of cognitive load. A regression equation predicted the percentage of Black celebrities chosen with cortisol reactivity (proxy for bias), cognitive load, and their interaction, after controlling for baseline cortisol. The Cortisol Reactivity × Cognitive Load interaction was significant, $t(53) = -2.26$, $p < .028$ (see Figure 4). Consistent with predictions, among those in the no-load condition, higher cortisol reactivity during the interracial interaction was related to choosing more Black over White celebrities. That is, the greater the bias, the greater the preferences for outgroup members in the no-load condition ($\beta = .45$, $p < .006$). However, when participants were under cognitive load, the greater the cortisol reactivity, the fewer Black celebrities were chosen ($\beta = -.17$, *ns*), though the effect was not significant.

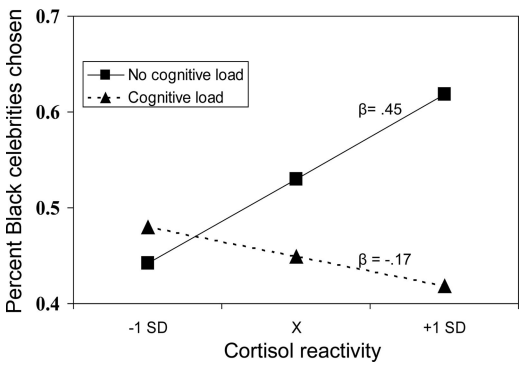

Figure 4. Experiment 3: Relationship between cortisol reactivity and percentage of preferred Black celebrities by cognitive load.

Evaluators' ratings of anxiety. We then turned to participants' manifested anxiety—the anxiety observed by the evaluators during the interview task. Male and female evaluators' ratings were highly correlated ($r = .49$, $p < .001$), so we combined the ratings into a single score of anxiety. This score was not significantly correlated with cortisol changes ($r = .10$) and thus represents a related but distinct measure of participants' anxiety, specifically anxiety as detected by the Black evaluators. Using anxiety ratings, cognitive load condition, and the interaction, we again predicted percentage of Black over White celebrities chosen. The interaction was significant, $t(50) = -2.42$, $p < .02$. Participants who were rated as appearing more anxious by the Black interviewers had chosen more Black celebrities, but only in the no-load condition ($\beta = .50$, $p < .032$). The load condition revealed the opposite pattern of findings: The more anxious participants were rated by the interviewers, the fewer Black celebrities were chosen, but this slope was not significant ($\beta = -.19$, $p < .27$).

General Discussion

This work explored whether majority group members engaged in more positive behavior toward stigmatized or minority group members relative to nonstigmatized or ingroup members. We argue that this overcorrection strategy requires self-regulatory effort and is based on the goal of appearing unprejudiced. Those who are most anxious in intergroup situations or more biased (measured implicitly or physiologically) were hypothesized as being most likely to display the overcorrection effect. Given the nature of self-regulation, we investigated whether overcorrection was fragile and would be attenuated when individuals' resources were taxed by cognitive load or stress. Taken together, these studies show that (a) when individuals have the resources to do so, they display positive biases toward stigmatized and minority group members relative to nonstigmatized or ingroup members; (b) the overcorrection effect appears to be threat based (i.e., the individuals who correct the most tend to be individuals who exhibit higher levels of racial bias or physiological signs of anxiety during social interactions with stigmatized or outgroup members); and (c) by reducing resources, either with stress or cognitive load, the overcorrection effect is attenuated. This last finding, especially, builds on previous work showing that compromised resources, due to cognitive load or alcohol consumption, can eliminate positive biases toward stigmatized members. Taken together, these studies expose the sometimes fragile nature of explicit outgroup preferences, and that these corrective processes may disappear when resources are exhausted.

It is important to note some limitations, particularly with regard to our measures of implicit bias, which are arguable flawed in different ways. The IAT as a measure of racial bias has its critics who question the measure's validity and predictive power (e.g., Blanton & Jaccard, 2006; Tetlock & Mitchell, 2009). Possibly more problematic is the use of physiological responses during interactions with outgroup partners and stigmatized others as measures of bias. Physiological changes of the neuroendocrine and cardiovascular systems can indicate a variety of changes in mental and nonmental states. Physiological reactivity during these social interactions are a function of individual differences, general anxiety, effort, motivation, and task-specific metabolic demands, such as changes in respiration required during speaking. So the changes

in reactivity during an interaction with stigmatized partners or outgroup evaluators reflect many affective states of which bias, linked to the specific partner with whom one is interacting, is just one component. A more complete design would be to have physiological reactions to ingroup and outgroup partners in a within-subjects design—though a within-subject design presents obstacles with regard to order effects, such that stressful speeches do not typically engender similar reactivity when completed a second time (Blascovich, Mendes, Vanman, & Dickerson, 2011). In any event, physiological responses cannot be conceived as isomorphically related to any affective or mental state (Mendes, 2009; in press); however, we suggest that these additional mental states add noise, which would reduce our ability to find relationships to choice and behavior like we observed here, and hence provide a conservative test of our hypotheses.

Is overcorrection a problem or a solution? We question whether overcorrection is disruptive in intergroup interactions or whether it leads to increased amiability between partners. One possibility is that overcorrection might result in a smoother interracial interaction. One study found that the White partners who Blacks perceived as most engaged in their interaction—and thus liked the most—were those with higher implicit race bias (Shelton, Richeson, Salvatore, & Trawalter, 2005). However, another possibility is that overcorrection might backfire if the positive treatment seems especially exaggerated or disingenuous. In one study, for example, Black participants who received positive social feedback from a White partner after a very brief interaction, exhibited cardiovascular threat responses, performed worse on a cognitive performance task, and showed more behavioral vigilance compared with Black participants who received negative feedback from White partners or Whites who received positive feedback from Blacks (Mendes et al., 2008). That is, there might be the belief that positive feedback, especially without justification, might be White individuals' clumsy attempts to correct for racial bias and ironically create more interracial tension. Unfortunately, when tension and stress increase, that may be exactly the situation where overcorrection breaks down.

Overcorrection has a cost, and it is ironically a cost that can make situations stressful. Trying to monitor one's behavior for signs of prejudice and correct for any outgroup bias that one possesses is likely to be a significant source of stress for majority group members in an intergroup interaction. White Americans monitor themselves for prejudiced behavior (Monteith, Deneen, & Tooman, 1996) and try to avoid behavior that would be offensive to African Americans (Norton et al., 2006). An individual's effort to correct for anti-Black bias could be related to the amount of anti-Black bias that he or she has. If those who are most biased against African Americans must put the most effort into controlling their prejudiced behavior during intergroup interactions, this may explain why it is those with the highest implicit anti-Black bias who are most cognitively depleted following an intergroup interaction (Richeson & Shelton, 2003). However, a desire to present oneself as unprejudiced can be threatening even in the absence of any underlying bias. Simply asking one group of participants to try to act in a nonprejudiced manner increased their anxiety during an intergroup interaction (Shelton, 2003).

Whether or not overcorrection is problematic when it occurs, it certainly appears to be problematic when it fails, and our studies suggest that overcorrection is fragile. Popular culture is replete with examples of failures of self-control leading to exposures of racial bias. In our studies, we found that conditions that taxed self-regulatory resources led to a reduction of overcorrection. If overcorrection is borne of a desire to appear unprejudiced, then it may be a weak strategy for achieving that goal, only useful when an individual has sufficient self-regulatory resources. Pursuing life experiences that could relieve anxiety around minority group members or change underlying attitudes toward them is likely to be a more resilient and permanent way to achieve a goal of egalitarian behavior.

References

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 256–274. doi:10.1037/0033-2909.111.2.256

Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology, 95,* 918–932. doi:10.1037/a0011990

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology, 90,* 272–287. doi:10.1037/0022-3514.90.2.272

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74,* 1252–1265. doi:10.1037/0022-3514.74.5.1252

Bergsieker, H. B., Shelton, J. N., & Richeson, J. A. (2010). To be liked versus respected: Divergent goals in interracial interactions. *Journal of Personality and Social Psychology, 99,* 248–264. doi:10.1037/a0018474

Bertrand, M., & Mullainathan, S. (2003). *Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination* (NBER Working Paper No. 9873). Cambridge, MA: National Bureau of Economic Research.

Blanton, H., & Jaccard, J. (2006). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 113,* 155–166. doi:10.1037/0033-295X.113.1.155

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of Personality and Social Psychology, 80,* 253–267. doi:10.1037/0022-3514.80.2.253

Blascovich, J., Mendes, W. B., & Seery, M. D. (2002). Intergroup encounters and threat: A multi-method approach. In D. M. Mackie & E. R. Smith (Eds.), *From prejudice to intergroup emotions: Differentiated reactions to social groups* (pp. 89–109). New York, NY: Psychology Press.

Blascovich, J., Vanman, E., Mendes, W. B., & Dickerson, S. (2011). *Social psychophysiology for social and personality psychology*. Los Angeles, CA: Sage.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. doi:10.1037/0022-3514.56.1.5

Devine, P. G., Monteith, M., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60,* 817–830. doi:10.1037/0022-3514.60.6.817

Devine, P. G., Plant, A. E., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82,* 835–848. doi:10.1037/0022-3514.82.5.835

Dovidio, J. F., Pearson, A. R., & Orr, P. (2008). Social psychology and neuroscience: Strange bedfellows or a healthy marriage? *Group Pro-*

cesses & Intergroup Relations, 11, 247–263. doi:10.1177/1368430207088041

Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology, 28,* 94–100. doi:10.1037/h0035582

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–517. doi:10.1037/0022-3514.60.4.509

Greenwald, A. G., McGhee, D. E., & Schwarz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480. doi:10.1037/0022-3514.74.6.1464

Harber, K. D. (1998). Feedback to minorities: Evidence of a positive bias. *Journal of Personality and Social Psychology, 74,* 622–628. doi:10.1037/0022-3514.74.3.622

Harber, K. D., Stafford, R., & Kennedy, K. A. (2010). The positive feedback bias as a response to self-image threat. *British Journal of Social Psychology, 49,* 207–218. doi:10.1348/014466609X473956

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *NeuroReport, 11,* 2351–2354. doi:10.1097/00001756-200008030-00004

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77,* 379–386. doi:10.1037/0022-3514.77.2.379

Mendes, W. B. (2009). Assessing the autonomic nervous system. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 118–147). New York, NY: Guilford Press.

Mendes, W. B. (in press). Inspired by the question not the measure: Exploiting neurobiological responses in the service of intergroup research. In N. Ellemers, D. Scheepers, & B. Derks (Eds.), *The nature of prejudice: A neuroscience perspective*. New York, NY: Psychology Press.

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interaction with White and Black men. *Personality and Social Psychology Bulletin, 28,* 939–952. doi:10.1177/01467202028007007

Mendes, W. B., Major, B., McCoy, S., & Blascovich, J. (2008). How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance. *Journal of Personality and Social Psychology, 94,* 278–291. doi:10.1037/0022-3514.94.2.278

Monteith, M. J., Deneen, N. E., & Tooman, G. D. (1996). The effect of social norm activation on the expression of opinions concerning gay men and Blacks. *Basic and Applied Social Psychology, 18,* 267–288. doi:10.1207/s15324834basp1803_2

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences, 33,* 417–433. doi:10.1017/S0140525X10000865

Norton, M. I., Sommers, S. R., Apfelbaum, E. P., Pura, N., & Ariely, D. (2006). Color blindness and interracial interaction: Playing the political correctness game. *Psychological Science, 17,* 949–953. doi:10.1111/j.1467-9280.2006.01810.x

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36–88. doi:10.1080/10463280701489053

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, E. A. (2005, July 29). The role of social groups in the persistence of learned fear. *Science, 309,* 785–787. doi:10.1126/science.1113551

Page-Gould, E., Mendoza-Denton, R., & Tropp, L. R. (2008). With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship. *Journal of Personality and Social Psychology, 95,* 1080–1094. doi:10.1037/0022-3514.95.5.1080

Plant, E. A., & Devine, P. G. (1998). Internal and external sources of motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75,* 811–832. doi:10.1037/0022-3514.75.3.811

Pryor, J. B., Reeder, G. D., Yeadon, C., & Hesson-McInnis, M. (2004). A dual-process model of reactions to perceived stigma. *Journal of Personality and Social Psychology, 87,* 436–452. doi:10.1037/0022-3514.87.4.436

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., & Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience, 6,* 1323–1328. doi:10.1038/nn1156

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive functioning. *Psychological Science, 14,* 287–290. doi:10.1111/1467-9280.03437

Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes & Intergroup Relations, 6,* 171–185. doi:10.1177/1368430203006002003

Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science, 16,* 397–402. doi:10.1111/j.0956-7976.2005.01547.x

Tetlock, P. E., & Mitchell, G. (2009). Implicit bias and accountability systems: What must organizations do to prevent discrimination? *Research in Organizational Behavior, 29,* 3–38. doi:10.1016/j.riob.2009.10.002

Trawalter, S., Richeson, J. A., & Shelton, J. N. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review, 13,* 243–268. doi:10.1177/1088868309345850

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73,* 941–959. doi:10.1037/0022-3514.73.5.941

Vorauer, J. D., & Turpie, C. A. (2004). Disruptive effects of vigilance on dominant group members' treatment of outgroup members: Choking versus shining under pressure. *Journal of Personality and Social Psychology, 87,* 384–399. doi:10.1037/0022-3514.87.3.384

Received February 25, 2012
Revision received July 11, 2012
Accepted July 11, 2012 ■